Solution structure and base pair opening kinetics of the i-motif dimer of d(5mCCTTTACC): a noncanonical structure with possible roles in chromosome stability

Sylvie Nonin, Anh Tuan Phan, Jean-Louis Leroy

Solution structure and base pair opening kinetics of the i-motif dimer of d(5mCCTTTACC): a noncanonical structure with possible roles in chromosome stability

Sylvie Nonin, Anh Tuan Phan and Jean-Louis Leroy*

Background: Repetitive cytosine-rich DNA sequences have been identified in telomeres and centromeres of eukaryotic chromosomes. These sequences play a role in maintaining chromosome stability during replication and may be involved in chromosome pairing during meiosis. The C-rich repeats can fold into an 'i-motif' structure, in which two parallel-stranded duplexes with hemiprotonated C·C$^+$ pairs are intercalated. Previous NMR studies of naturally occurring repeats have produced poor NMR spectra. This led us to investigate oligonucleotides, based on natural sequences, to produce higher quality spectra and thus provide further information as to the structure and possible biological function of the i-motif.

Results: NMR spectroscopy has shown that d(5mCCTTTACC) forms an i-motif dimer of symmetry-related and intercalated folded strands. The high-definition structure is computed on the basis of the build-up rates of 29 intraresidue and 35 interresidue nuclear Overhauser effect (NOE) connectivities. The i-motif core includes intercalated interstrand C·C$^+$ pairs stacked in the order 2*·8 / 1·7* / 1*·7 / 2·8* (where one strand is distinguished by an asterisk and the numbers relate to the base positions within the repeat). The TTTA sequences form two loops which span the two wide grooves on opposite sides of the i-motif core; the i-motif core is extended at both ends by the stacking of A6 onto C2·C8$^+$. The lifetimes of pairs C2·C8$^+$ and 5mC1·C7$^+$ are 1 ms and 1s, respectively, at 15°C. Anomalous exchange properties of the T3 imino proton indicate hydrogen bonding to A6 N7 via a water bridge. The d(5mCCTTTTCC) deoxyoligonucleotide, in which position 6 is occupied by a thymidine instead of an adenine, also forms a symmetric i-motif dimer. However, in this structure the two TTTT loops are located on the same side of the i-motif core and the C·C$^+$ pairs are formed by equivalent cytidines stacked in the order 8*·8 / 1·1* / 7*·7 / 2·2*.

Conclusions: Oligodeoxynucleotides containing two C-rich repeats can fold and dimerize into an i-motif. The change of folding topology resulting from the substitution of a single nucleoside emphasizes the influence of the loop residues on the i-motif structure formed by two folded strands.

Address: Groupe de Biophysique, de l'Ecole Polytechnique et de l'URA, D-1254 du CNRS, 91128 Palaiseau, France.

*Corresponding author.
E-mail: jll@pmc.polytechnique.fr

Key words: hydration, i-motif, proton exchange, solution structure, telomere

Received: **8 May 1997**
Revisions requested: **20 June 1997**
Revisions received: **17 July 1997**
Accepted: **17 July 1997**

Structure 15 September 1997, **5**:1231–1246

Introduction

Telomeres, the ends of eukaryotic chromosomes, include a DNA strand of short cytosine-rich repeats complementary to the repeats of the corresponding guanosine-rich strand. These sequences play a role in maintaining chromosome integrity and stability during replication. X-ray crystallography [1] and NMR [2,3] spectroscopy have established that the G-rich repeats can fold *in vitro* into a four-stranded structure of stacked, planar G tetrads. It has been suggested that such tetrads could be involved in the association of chromatids during mitosis and meiosis, and in the regulation of the telomerase activity [4].

The complementary C-rich strand can also adopt a four-stranded arrangement [5,6]. Protonation at the cytidine N3 position allows the formation of parallel duplexes whose strands are held together by hemiprotonated C·C$^+$ pairs [7,8]. The intercalation in a head-to-tail orientation of two duplexes containing C-rich repeats results in the formation of a four-stranded structure called the 'i-motif'.

The i-motif may be formed by the intermolecular assembly of four C-rich oligonucleotides, each containing one stretch of cytidines [9–13], by two oligonucleotides each containing two cytidine stretches [5,14], or by the intramolecular folding of one oligonucleotide containing four cytidine stretches [6,15] (X Han *et al.*, unpublished data].

Deoxyoligonucleotides with one cytidine stretch generally form a single tetrameric structure. This is the case for

d(TCC), d(5mCCT) [13], d(5mCCTCC) [16], d(5mCCC), d(T5mCCC) (J-LL, unpublished data) and d(TCCCCC) [9,17]. The NMR spectra of a single deoxyoligonucleotide strand containing four repeats of the telomeric sequence of vertebrate d(CCCTAA) and *Tetrahymena* d(CCCCAA) [6] display the NMR determinants of the i-motif, but the presence of more than one spin system per residue discloses the existence of multiple conformations. Similarly, the poor resolution of the NMR spectrum of an oligonucleotide strand with four human centromeric repeats (CCATT) suggests conformational exchange between several i-motif structures (SN and J-LL, unpublished data).

The length and composition of the loops connecting the cytidine stretches may influence the folding topology and stability of the i-motif. To investigate this problem, we have studied several i-motif forming sequences, in which the spacers between the cytidine stretches differ from the natural ones.

In this work, we solve the solution structure of the i-motif dimer of d(5mCCTTTACC). Cytidine methylation is not a prerequisite to the i-motif formation, but a 5-methyl cytidine (5mC) at the 5′-end improves the spectral dispersion [13]. The well-resolved NMR spectrum shows that this oligonucleotide forms a symmetrical dimer at slightly acidic pH and a high-definition structure is obtained. The dimer has a core of four intercalated C·C$^+$ pairs and the TTTA spacers loop on each side of the core. Imino proton exchange time measurements provide the base pair lifetimes and give evidence for a water bridge in the loops between T3 H3 and A6 N7.

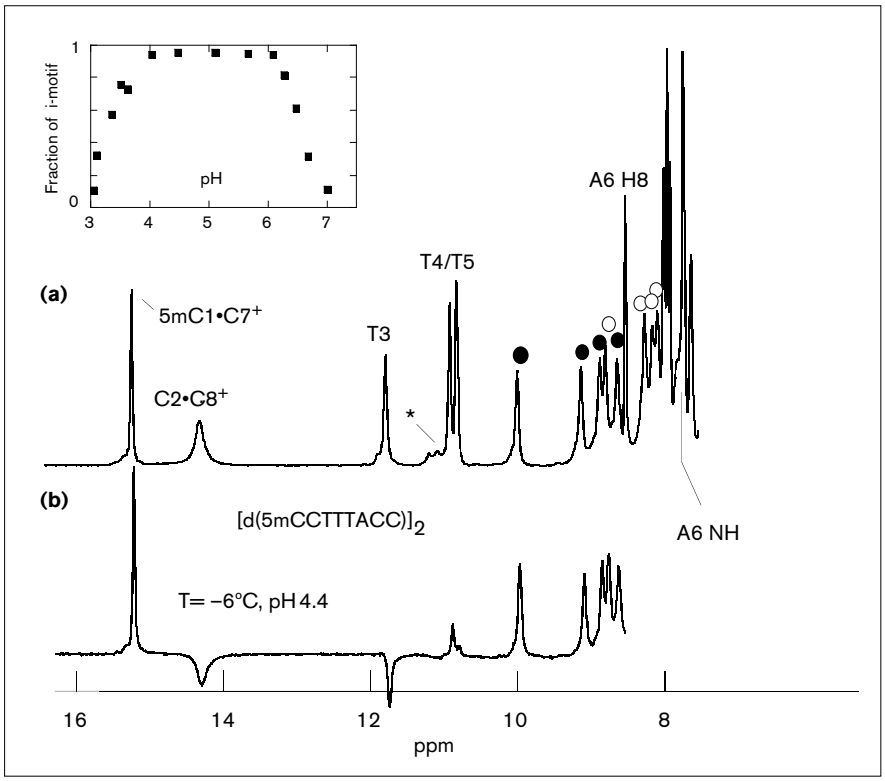
The d(5mCCTTTTCC) deoxyoligonucleotide also forms an i-motif dimer, but the structure is different: the two spacer loops are positioned close together on the same side of the i-motif core. The comparison of the two structures suggests that the stacking of adenosine A6 on the adjacent C·C$^+$ pair contributes to the stability of the $[d(5mCCTTTACC)]_2$ dimer.

Results

Stoichiometry

The NMR spectrum of d(5mCCTTTACC) shows two sets of peaks corresponding to two species in slow exchange on the NMR time scale. The relative intensity of the two sets of peaks depends on pH, temperature and concentration. The spectrum which dominates at low concentration and high temperature is ascribed to the single strand, the other to a multimer (Figure 1). Based on the relative intensities of the adenosine H8 resonances, at 8.39 ppm for the multimer and 8.45 ppm for the single strand, the multimer concentration increases as the square of the single strand concentration (Figure 2): hence, it is a dimer. The dissociation

Figure 1



(a) Exchangeable, aromatic and H1′ proton spectrum of $[d(5mCCTTTACC)]_2$. The presence of two cytidine imino protons (15.47 and 14.52 ppm), four cytidine external (○) and four internal (●) amino proton peaks is evidence for two nonsymmetrical hemiprotonated C·C$^+$ pairs. The thymidine imino proton peaks of the small fraction of single-stranded d(5mCCTTTACC) are labeled by an asterisk. **(b)** The spectrum recorded 50 ms after selective inversion of the water magnetization shows the fast exchange rate of the 14.52 ppm C imino proton and of the 11.6 ppm T imino. Inset: the dimer fraction, as determined from the area of the 15.47 ppm imino proton peak, is displayed versus pH. Solution conditions: H_2O 90%, pH 4.4, T = −6°C, strand concentration 8 mM.

constant is 3.5×10^{-4} M at 21°C. For a strand concentration of 8 mM, the dimer is the main species between pH 3.3 and pH 6.6 at 0°C (inset of Figure 1).

The stoichiometry was confirmed by gel filtration chromatography: the rate of gel filtration corresponds to that of a dimer at a strand concentration of 5 mM, and to that of a monomer at a concentration of 5×10^{-5} M.

Proton spectra and symmetry elements

Exchangeable protons

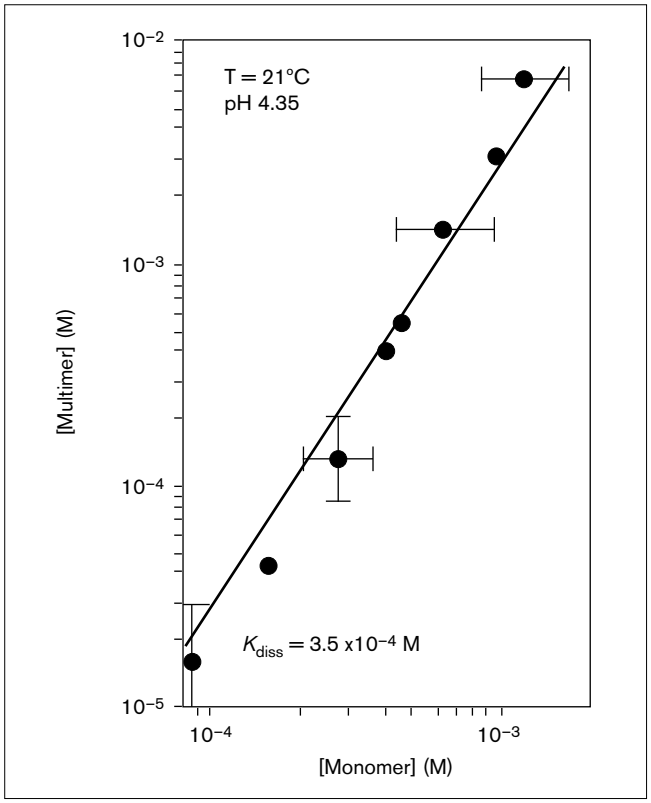
At –6°C, the proton spectrum exhibits two cytidine imino proton peaks (15.47 and 14.52 ppm) and four couples of amino proton peaks, disclosing the formation of two hemiprotonated C·C$^+$ pairs (Figure 1). Each imino proton is nuclear Overhauser effect (NOE)-connected to two distinct amino proton groups (Figure 3). This shows that the two cytidines of each C·C$^+$ pair are not equivalent. Three thymidine imino proton peaks are detected between 10.8 and 12 ppm. All the cytidine and thymidine imino peaks have the same areas. The rate of exchange with water of the 11.90 ppm thymidine imino proton is particularly fast, as shown by the spectrum recorded 50 ms after selective inversion of water (Figure 1). The adenosine amino proton peaks are identified by a strong NOE spectroscopy (NOESY) cross-peak between two exchangeable protons (cross-peak e6, Figure 3d). The observation of two distinct adenosine amino proton peaks shows that rotation around C6–N6 is much slower than in the monomer. The downfield shift of the resolved A6 amino proton at 6.86 ppm in a pH range close to the adenosine pK, p$K_{(A,N1)} = 4.05$ at 0°C , is attributed to protonation at the N1 position.

Dimer symmetry and folding

The total correlation spectroscopy (TOCSY) and NOESY spectra display only eight spin systems (Figures 3 and 4). This shows that the two strands are identical on the NMR time scale and that the dimer has a twofold symmetry axis. The three thymidine and the three cytidine spin systems were readily identified by the methyl–H6 and H5–H6 cross-peaks, respectively. A6 was identified by the non-exchangeable peak at 8.39 ppm (H8) and by the slow longitudinal relaxation rate of H2. The 5mC1 residue was identified by the strong intraresidue NOEs between the amino protons and the methyl group (cross-peaks f1 and f′1, Figure 3b). The spin system of each residue was identified (except for H5′ and H5″ protons) by the standard patterns of intraresidue NOESY and TOCSY connectivities (Table 1).

Characteristic NOESY cross-peaks provide evidence for an i-motif structure: interresidue cytidine amino protons–H2′/H2″ (δ, χ, ε and ϕ, Figure 3b) and H1′–H1′ cross-peaks (α and β, Figure 4d); and the strong intraresidue H3′–H6 cross-peaks.

Figure 2



NMR titration of the multimer versus monomer concentration of d(5mCCTTTACC), as determined by the areas of their A6 H8 peaks. A line of slope 2, corresponding to the formation of a dimer whose dissociation constant is 3.5×10^{-4} M, is drawn through the data points. The experimental conditions were D_2O 99.9%, potassium phosphate 20 mM, pH 4.35, T = 21°C.

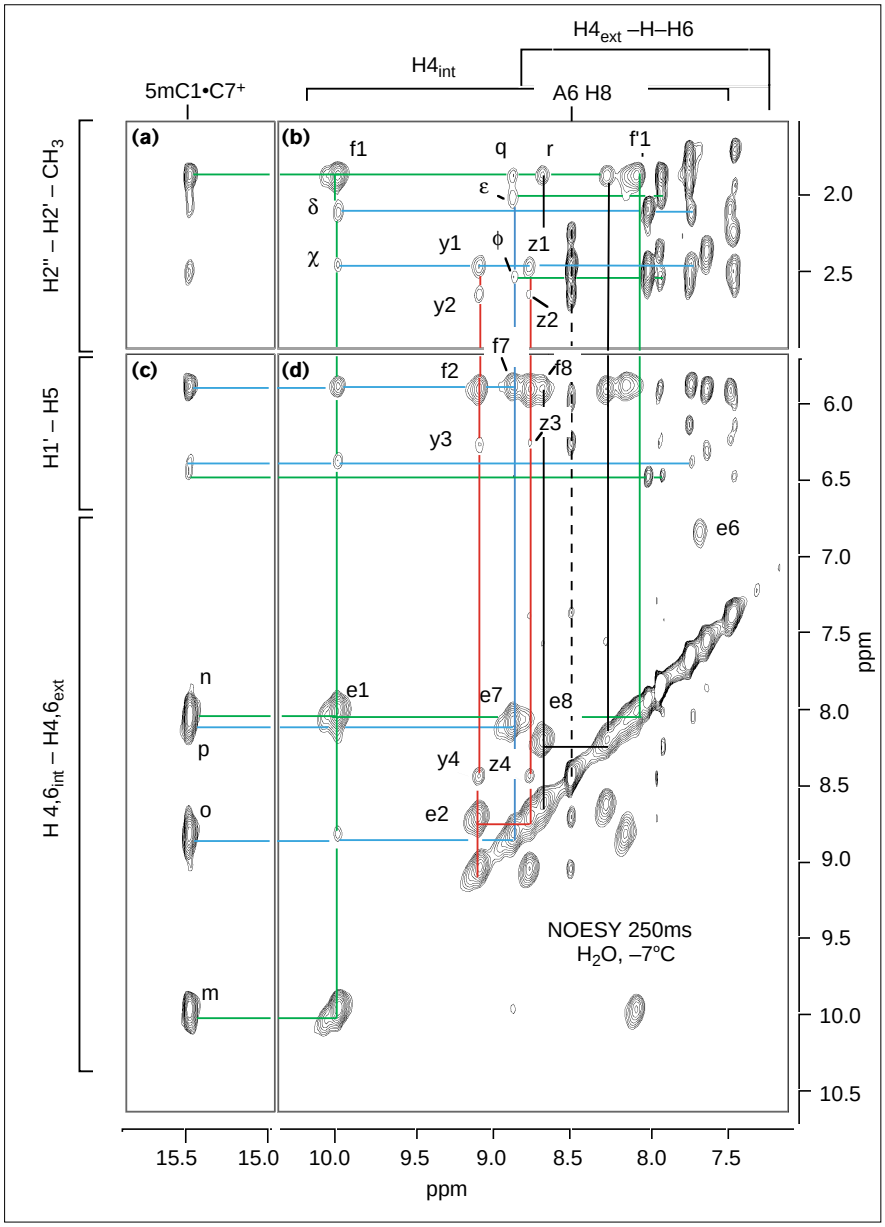
Three families of structures can account for the formation of a symmetrical i-motif dimer (Figure 5). Scheme (a) is excluded because its paired cytidines are equivalent (see above). Schemes (b) and (c) have the same i-motif core formed by two identical 5mC1·C7$^+$ and two identical C2·C8$^+$ pairs. The C·C$^+$ pairs are intrastrand in structure (b) and interstrand in structure (c).

The i-motif core: proton assignment and intercalation topology

The assignment of the spin systems was derived from the NOESY cross-peaks characteristic of the i-motif (see Materials and methods section). Due to the lack of sequential connectivities within the i-motif, the proton assignment procedure and the determination of the intercalation topology proceeded concurrently.

The NOEs between 5mC1 amino protons and the 15.47 ppm imino proton (cross-peaks m and n, Figure 3c) identify the imino proton of the 5mC1·C7$^+$ pair which is in

Figure 3



NOESY spectrum of [d(5mCCTTTACC)]$_2$ in H_2O at 250 ms mixing time. **(a,b)** Intraresidue cross-peaks between the internal/external amino protons and CH_3/H5 protons are labeled by letters 'f' and 'f'', respectively, followed by the residue number. The amino–H2′/H2″ cross-peaks are displayed in (b): δ, 5mC1 $H4_{int}$–C7 H2′; χ, 5mC1 $H4_{int}$–C7 H2″; ε, C7 $H4_{int}$–5mC1 H2′; ϕ, C7 $H4_{int}$–5mC1 H2″ and $H4_{ext}$–H2′/H2″. Interresidue connectivities between C2 $H4_{int}$/$H4_{ext}$ and A6 protons are designated y1 to y4, and z1 to z4, respectively (b,d): y1/z1, C2 $H4_{int}$/$H4_{ext}$–A6 H2′; y2/z2, C2 $H4_{int}$/$H4_{ext}$–A6 H2″; y3/z3, C2 $H4_{int}$/$H4_{ext}$–A6 H1′; y4/z4, C2 $H4_{int}$/$H4_{ext}$–A6 H8. Cross-peaks q and r are assigned as follows: q, C7 $H4_{int}$–5mC1 CH_3; r, C8 $H4_{ext}$–5mC1 CH_3. **(c)** The cross-peaks between the imino proton at 15.47 ppm and the amino protons of 5mC1 (cross-peaks m and n) and C7 (cross-peaks o and p) establish that the C·C$^+$ pairs are not symmetrical. **(d)** The intraresidue cross-peaks between internal and external amino protons are labeled 'e' followed by the residue number. Solution conditions: 90% H_2O–10% D_2O, pH 4.6, T = –7°C. The blue and the green lines trace cross-peaks involving 5mC1 and C7 residues, respectively. The red and the black lines trace some connectivities involving C2 and C8, respectively. Spectrometer frequency 360 MHz.

turn connected to the amino protons of C7 (cross-peaks o and p, Figure 3c). The reciprocal amino–H2′/H2″ cross-peaks connecting 5mC1 and C7 (cross-peaks δ and χ between the internal amino proton of 5mC1 and C7 H2′/H2″, and cross-peaks ε and ϕ between the internal amino proton of C7 and 5mC1 H2′/H2″, Figure 3b) characterize the intercalation topology of the i-motif (Figure 6). The distances derived from their build-up (2.86 Å for 5mC1 $H4_{int}$–C7 H2′ and 3.43 Å for C7 $H4_{int}$–5mC1 H2′) are too short for base-paired residues whose corresponding distances are larger than 5.5 Å. These NOEs disclose the short interproton distances between the two adjacent face-to-face opposed 5mC1·C7$^+$ pairs. The observation of amino–H2′/H2″ cross-peaks between 5mC1 and C7 implies that the 5mC1–C2 and C7–C8 stretches run in opposite directions across the wide grooves. The face-to-face intercalation of two 5mC1·C7$^+$ pairs brings close together the methyl group of 5mC1 and C7 H5 which are connected by a strong cross-peak (not shown).

The 14.52 ppm imino proton peak is assigned to the imino proton of pair C2·C8$^+$. Its fast exchange rate (Figure 1) is consistent with the external position of this pair, relative to 5mC1·C7$^+$. This leads to the stacking order 2·8 / 1·7 / 1·7 / 2·8, in which the cytidines at the 3′-end position of each C–C stretch occupy the external

positions of the i-motif core. At this stage, we still ignore if the pairs are intra- or interstrand.

The C2 and C8 spin systems were identified by the H1′–H1′ cross-peaks with 5mC1 and C7, respectively (cross-peaks α and β, Figure 4d). In addition, 5mC1 H6 shows a weak cross-peak with C2 H1′ (not shown), and the methyl group of 5mC1 is NOE-connected to the internal amino proton of C8 (cross-peak r, Figure 3b).

The TTTA stretch: proton assignment and folding

Proton assignment in the TTTA stretch starts from A6 H8. The orientation of the A6 ring with respect to C2·C8$^+$ is determined by the NOEs between A6 H2′/H2″/H1′/H8 and C2 amino protons (cross-peaks y1–z1, y2–z2, y3–z3 and y4–z4, respectively, Figures 3b,d). Sequential connectivities in the aromatic–H1′, H3′ and H2′/ H2″ regions can be followed from A6 to T4 (Figures 4 and 6). Direct NOE cross-peaks between A6 H8 and T5 H6, between T5 H6 and T4 H6, and between T5 CH3 and T4 H6 (cross-peak x1, Figure 4a), indicate a stacked arrangement of A6, T5 and T4 rings.

The last thymidine, T3, is identified by six direct NOEs with C2 (Figure 6; cross-peaks x2 and x4, Figure 4). The build-up rate of the weak NOE cross-peak between T3 H3 and A6 H8 corresponds to a distance of 3.8 ± 0.4 Å. This cross-peak and those detected between T3 sugar protons (H4′ and H1′) and T5 ring protons (cross-peak x5, Figures 4c and 6) show that the TTTA stretch forms a loop. There are no NOE connectivities at short mixing times between T3 and T4.

The structural family with intrastrand base pairs (Figure 5c) cannot account simultaneously for the cross-peaks observed between A6, T3 and C2·C8$^+$, and those connecting the residues along the TTTA stretches. This leaves us with the structural family of Figure 5b, whose base pairs connect the 5mC1–C2 and C7–C8 stretches of two distinct strands. This implies that the antiparallel 5mC1–C2 and C7–C8 stretches separated by the wide grooves belong to the same strand and therefore that the TTTA loops span the wide groove.

Proton exchange

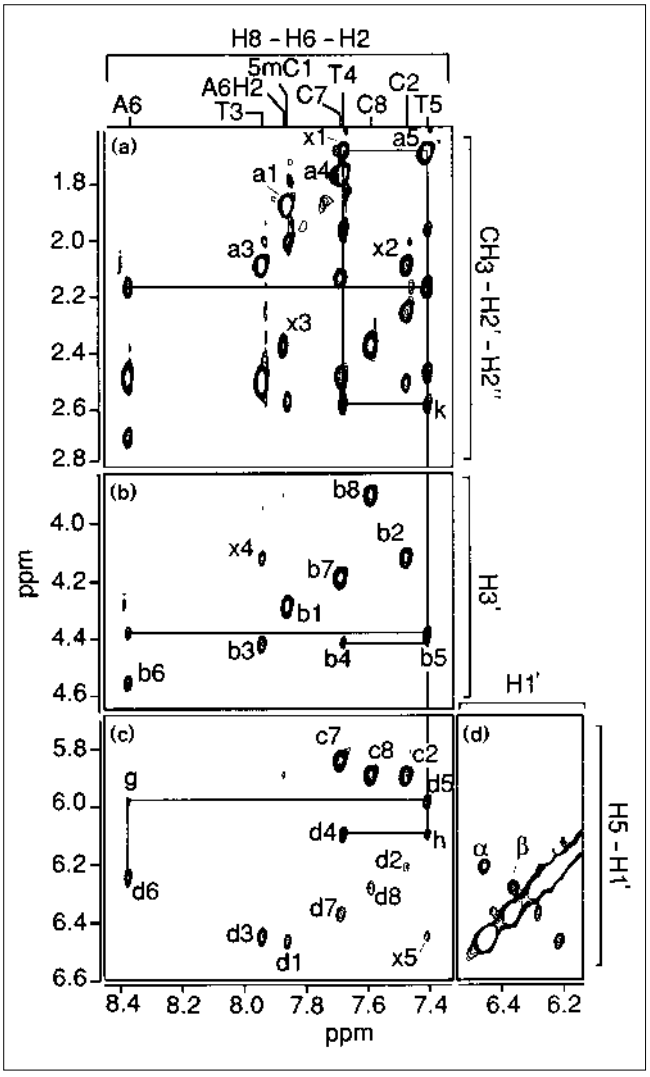
Cytidine imino proton exchange

At 15°C, imino proton exchange from 5mC1·C7$^+$ is one thousand times slower than from C2·C8$^+$ (1 s and 1 ms, respectively). The activation energies (Figure 7a) are 100 and 60 kJ M^{-1}, respectively, for 5mC1·C7$^+$ and C2·C8$^+$. As always in C·C$^+$ pairs, imino proton exchange is not affected upon addition of proton acceptors (e.g. $H_2PO_4^-$, 0.1 M, pH 6) and does not depend on pH.

Thymidine imino proton exchange

The similarity of the rates of exchange catalysis by hydroxide (Figure 7b) and phosphate (Figure 7d) on the monomer and on T3, T4 and T5 shows that the imino protons of the thymidines of the TTTA loop are accessible to the solvent. Between pH 3.5 and pH 7, imino proton exchange from T3 is faster than from the

Figure 4



NOESY of [d(5mCCTTTACC)]$_2$ in D_2O at 250 ms mixing time. NOE cross-peaks between aromatic protons and: CH_3/H2′/H2″ protons **(a)**, H3′ protons **(b)**, H5 and H1′ protons **(c)**. The characteristic H1′–H1′ cross-peaks are displayed in **(d)**: α, 5mC1 H1′–C2 H1′; β, C7 H1′–C8 H1′. Intraresidue cross-peaks are designated by letters followed by the residue number: a, (H6–CH_3); b, (H6/H8–H3′); c, (H5–H6) and d, (H6/H8–H1′). The NOE connectivities between aromatic protons and their own and 5′-flanking H1′, H3′ and H2′ or H2″ protons are traced from A6 to T4. Sugar–base interresidue cross-peaks in the TTTA loop are labeled: g, (A6 H8–T5 H1′); h, (T5 H6–T4 H1′); i, (A6 H8–T5 H3′); j, (A6 H8–T5 H2′); and k, (T5 H6–T4 H2″). Cross-peak x1 is assigned to (T5 CH_3–T4 H6), x2 to (C2 H6–T3 CH_3), x3 to (C1 H6–C8 H2′/H2″), x4 to (T3 H6–C2 H3′) and x5 to (T5 H6–T3 H1′). Solution conditions: 99.99 % D_2O, pH 4.6, T = 7°C; spectrometer frequency 600 MHz.

Table 1

Proton chemical shifts (ppm) of $[d(5mCCTTTACC)]_2$.*

Base	H3	Internal/external amino protons	H6/H8	H5/H2/CH3	H1′	H3′	H4′	H2″	H2′
5mC1	15.47	10.03/8.11	7.84	1.99	6.43	4.76	4.04	2.46	1.91
C2	14.52	9.10/8.76	7.44	7.15	6.17	4.59	4.39	2.39	2.15
T3	11.89		7.92		6.42	4.90	4.41	2.39	2.39
T4	10.99 or 10.88		7.65	1.71	6.06	4.88	4.24	2.87	2.47
T5	10.88 or 10.99		7.37	1.77	5.99	4.86		2.43	2.12
A6		7.61/6.86	8.39	7.86	6.20	5.03	4.30	2.59	2.38
C7	15.47	8.85/8.24	7.65	5.80	6.36	4.66	4.22	2.04	2.38
C8	14.52	8.62/8.30	7.56	5.86	6.24	4.37		2.28	2.28

*Imino and amino protons (–7°C) and non-exchangeable protons (10°C) at pH 4.6.

monomer, indicating exchange catalysis by an intrinsic proton acceptor (Figure 7c).

Structure analysis

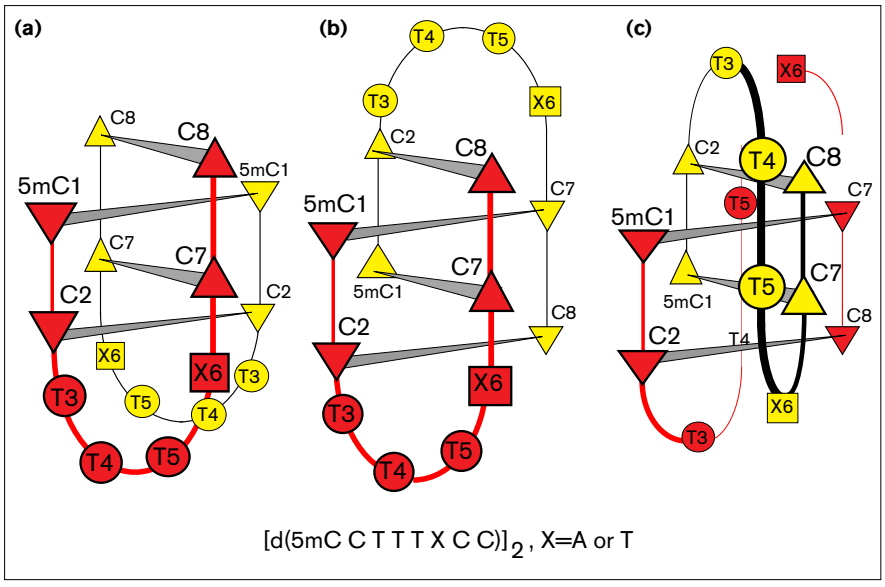
The structure was computed using the simulated-annealing method [18] (as described in the Materials and methods section; the interresidue distance constraints are displayed in Figure 6). Figure 8 shows the distribution of the constraints used for each residue in the computation.

We analysed the structure using 14 conformers selected for their low NOE-related energy. All of the conformers have the same intercalation topology and loop structure. The root mean square deviation (rmsd), computed from the pair-wise comparison of the conformers aligned on the heavy atoms of the bases, is plotted for each residue in Figure 8. The deviations from ideal geometry of the conformer of lowest energy and the structure parameters are shown in Tables 2 and 3, respectively.

The i-motif core

The core of the dimer is an i-motif formed of two strands connected by four C·C$^+$ pairs and related by a twofold symmetry. The stacking order is 2*·8 / 1·7* / 1*·7 / 2·8* (where one strand is labelled by an asterisk). The C1′–C1′ distances in pairs C2·C8$^+$ and 5mC1·C7$^+$ are 10.3 ± 0.5 Å and 9.1 ± 0.2 Å, respectively. The structure has two identical wide grooves and two different narrow grooves

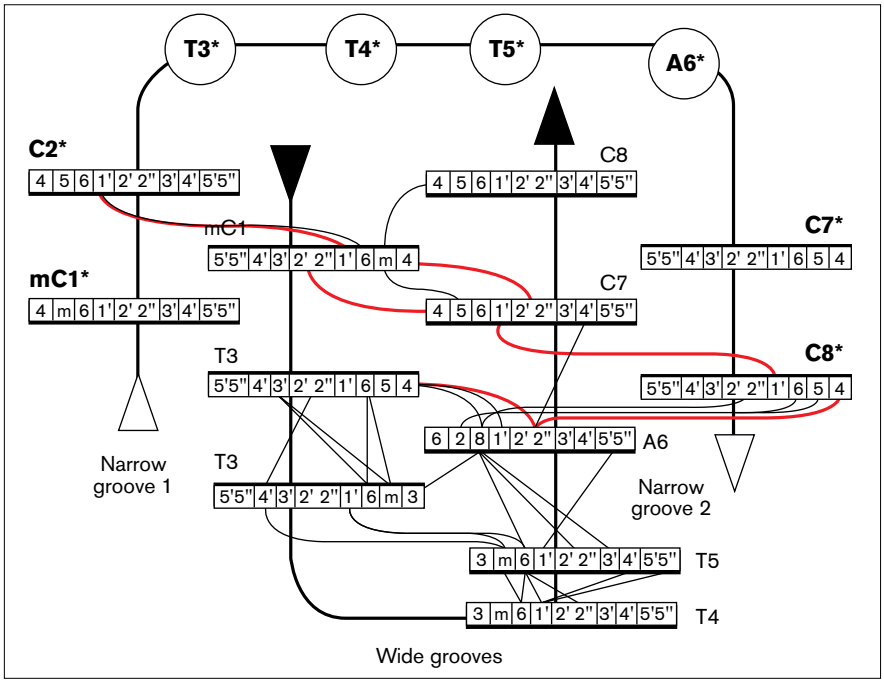
Figure 5



Schematic representation of the folding topologies of i-motif dimers **(a–c)** formed by two identical d(5mCCTTTXCC) strands. (where X is either A or T). The cytidines are drawn as triangles, and the thymidines T3, T4 and T5 as circles. Base paired cytidines are connected by gray arrows. One strand is shown in red, the other in yellow. In model (a), the loops are on the same side of the structure and can span the wide groove (as shown in the figure) or the narrow groove. The strands of models (a) and (b) are held together by interstrand C·C$^+$ pairs, and those of model (c) by intrastrand C·C$^+$ pairs. In model (c) the loops can occupy the narrow groove (as shown in the figure) or the wide groove. For each model, there is two possible intercalations: the external C·C$^+$ pair may be formed by the cytidines at the 5′-end or at the 3′-end of the CC stretches; the loops can also span either the narrow or the wide grooves. The observation that the C·C$^+$ pairs of the d(5mCCTTTACC) structure are formed of non-equivalent cytidines eliminates scheme (a). The NOE cross-peaks between the residues of the TTTA stretch, between C2 and A6, C2 and T3, and between C8 and A6, are incompatible with scheme (c) and validate scheme (b). $[d(5mCCTTTTCC)]_2$, whose C·C$^+$ pairs are formed of equivalent cytidines, adopts the folding topology of scheme (a).

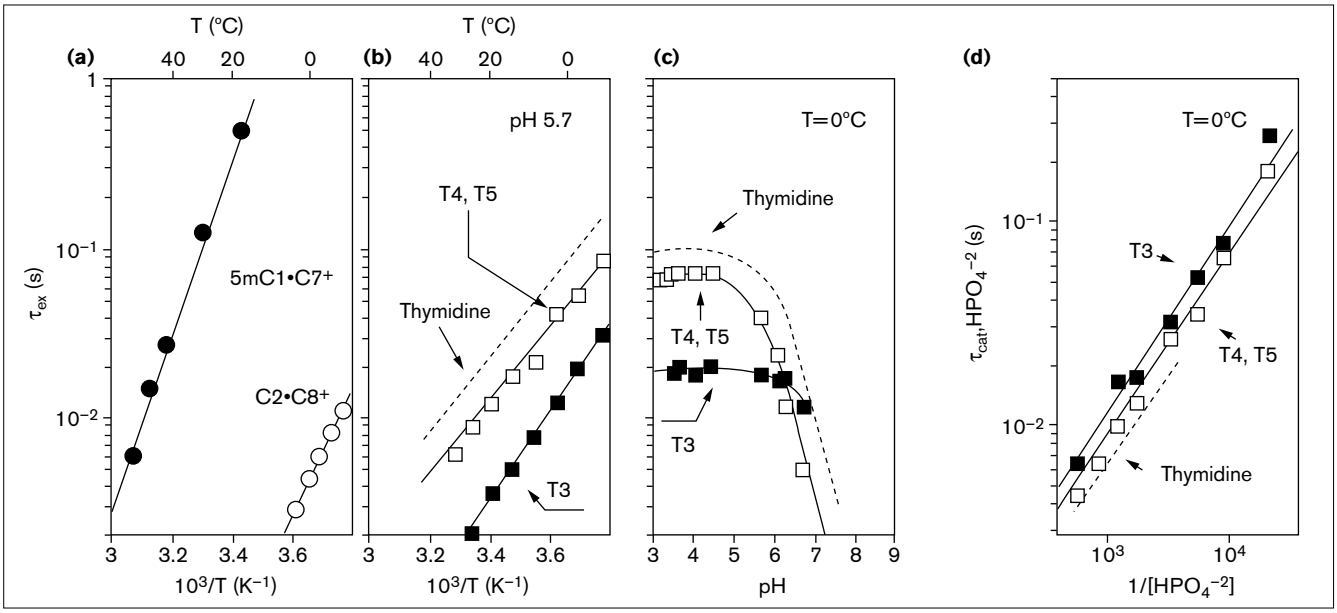
Figure 6

Schematic representation of the interresidue distance constraints used in the calculation of the [d(5mCCTTTACC)]$_2$ structure. As a consequence of the symmetry, the interresidue NOE connectivities provide two identical sets of distance constraints. For clarity, only one set is displayed and the constraints corresponding to interproton distances larger than 4.2 Å are omitted. The short distances between H1′ protons and between amino–H2′/H2″ protons, which constitute the fingerprints of the i-motif and characterize its intercalation topology, are drawn as heavy red lines. The lack of connectivities between T3 and T4 indicates a sharp turn in the TTTA loop. The C2–T3, A6–T5 and T5–T4 connectivities between ring protons indicate stacking of these residues.



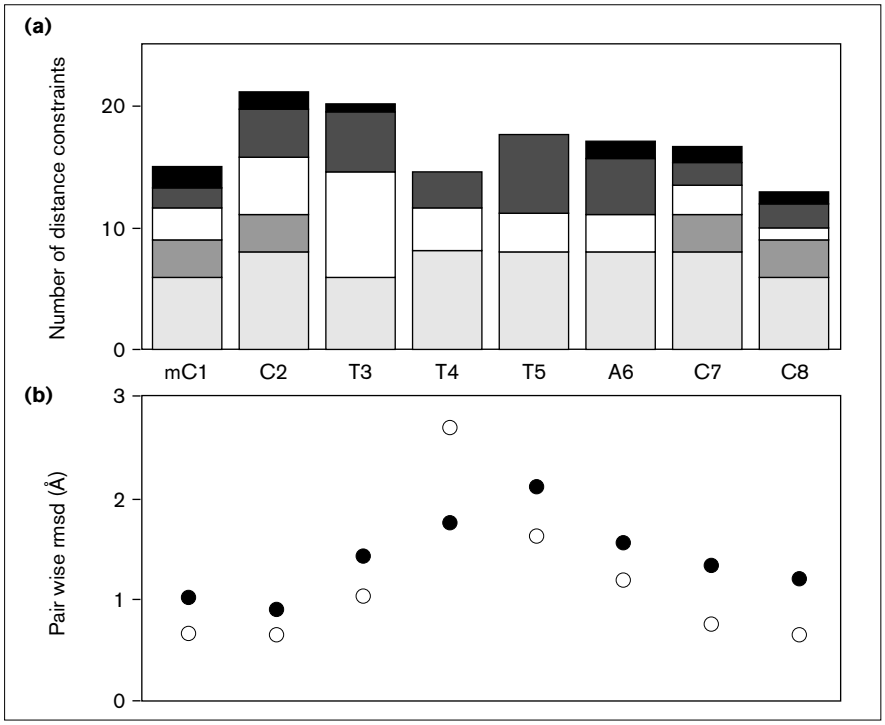
(Table 3). The schematic representation (Figure 9) shows that the twofold symmetry axis crosses the minor grooves midway between the two 5mC1·C7$^+$ pairs. The small buckle and propeller twist values (Table 3) indicate that the cytidine pairs are remarkably flat. The reduced overlap of pairs C2·C8$^+$ and 5mC1·C7$^+$ (Figure 10) allows

Figure 7



Imino proton exchange in [d(5mCCTTTACC)]$_2$. **(a)** Cytidine imino proton exchange times versus temperature: C2·C8$^+$ (○) and 5mC1·C7$^+$(●). **(b,c)** Thymidine imino proton exchange versus temperature at pH 5.7 and versus pH at 0°C: T3 (■), T4 and T5 (□), and the thymidine monomer (dashed line). **(d)** Phosphate catalysis of imino proton exchange at 0°C. The exchange times measured in the presence of phosphate for the imino proton of thymidine (dashed line), T3 (■), T4 and T5 (□) are comparable.

Figure 8



(a) Distribution of the distance constraints used in the structure calculation of $[d(5mCCTTTACC)]_2$. The constraints have been classified as follows: interresidue distances involving an exchangeable proton (black) and two non-exchangeable protons (darkest gray); hydrogen-bond lengths used to define the C·C$^+$ pairing (medium gray); intraresidue distance constraints (light gray); and repulsive constraints reflecting the absence of direct NOE cross-peaks at long mixing times (white). The intraresidue cross-peaks are counted as one, and the interresidue constraints as one half for each partner. **(b)** Pair wise root mean square deviation (rmsd; Å) in the heavy-atom coordinate positions of the bases (○) and sugar phosphate backbones (●) calculated for each residue from the coordinates of the 14 lower-energy conformers aligned on the heavy atoms of the cytidine rings.

an unusually small stacking interval (2.7 ± 0.6 Å) between these base pairs. The helical twist (18° ± 6) is typical for the i-motif. The bases are *anti*, the γ torsion angles are in the *trans* and *gauche*$^-$ domains, and the sugar puckers are in the N conformational range (Table 3).

The TTTA stretches

The TTTA stretches form two loops, each of which bridges one major groove on either side of the i-motif core which is extended by the stacking of A6 onto the external C2·C8$^+$ base pair (Figure 10). The T3 C1′–A6 C1′ distance (7.9 ± 0.4 Å) is shorter than that expected for a Watson–Crick base pair (10.4 Å), but close to the 8.6 Å of a Hoogsteen pair. However, the NOE-derived distance between T3 H3 and A6 H8 (3.8 ± 0.4 Å) is much longer than the corresponding distance in a Hoogsteen pair (2.7 Å). The aromatic rings of A6 and T3 are almost perpendicular to each other. Modeling performed with the T3 H3–A6 N7 distance constrained to a value compatible with hydrogen bonding, and enforcement of the planarity of the A6–T3 pair, results in strong violations of the NMR-derived NOE distance constraints. All the bases of the TTTA loop are *anti* with χ angles ranging from –106 ± 5° to –140 ± 5°. The sugar pucker of A6 is in the N conformational range and those of the thymidines are in the S conformation.

Table 2

Violations and deviations from ideal geometry in structural computations.

Number of NOE violations larger than 0.2 (Å)	4
Largest NOE violation (Å)	0.24
Rms violation of input distance constraints (Å)	0.06
Number of ideal bond angle violations larger than 6°	4
Largest ideal bond angle violation (°)	7.8
Rmsd from ideal bond angle (°)	1.3
Number of improper violations larger than 5°	0
Rmsd from ideal improper (°)	0.34
Bond violations larger than 0.05 Å	0
Rmsd from ideal bond lengths (Å)	0.006

Hydration sites

Following Otting *et al.* [19], we investigated the presence of long lived hydration water molecules within the structure by one-dimensional (1D)-NOE and 1D-ROE experiments. The signs of the ROE and NOE cross-peaks between the water peak and A6 H8, A6 H2 and 5mC1 H6, indicate the presence of hydration water molecules whose residence times are longer than 0.5 ns (Figure 11). The cross-peak intensities decrease as temperature increases, as expected if the cross-peaks arise by dipolar interactions with water. The intensity of the H_2O–A6 H8 cross-peak is comparable to that measured between water and A6 H2 in $d(CGCGAATTCGCG)_2$, a duplex considered as a reference for hydration studies of nucleic acids [20–22].

Table 3

Geometrical parameters of [d(5mCCTTTACC)]$_2$.*

Base	χ(°)	P(°)	Buckle (°)†	Propeller twist (°)‡	Step	Helical twist (°)
5mC1	−140 ± 5	25 ± 6	−1 ± 2	−1 ± 2	5mC1–C2	18 ± 6
C2	−128 ± 4	12 ± 9	−3 ± 2	−6 ± 2	C2–T3	56 ± 5
T3	−127 ± 5	155 ± 5	–	–	T3–T4	−115 ± 10
T4	−108 ± 5	188 ± 6	–	–	T4–T5	35 ± 5
T5	−136 ± 4	98 ± 30	–	–	T5–A6	−7 ± 13
A6	−106 ± 5	−23 ± 21	−17 ± 6	−19 ± 7	A6–C7	70 ± 7
C7	−118 ± 4	−33 ± 12	−16 ± 4	−3 ± 3	C7–C8	19 ± 6
C8	−123 ± 5	11 ± 30	−14 ± 4	0 ± 4		

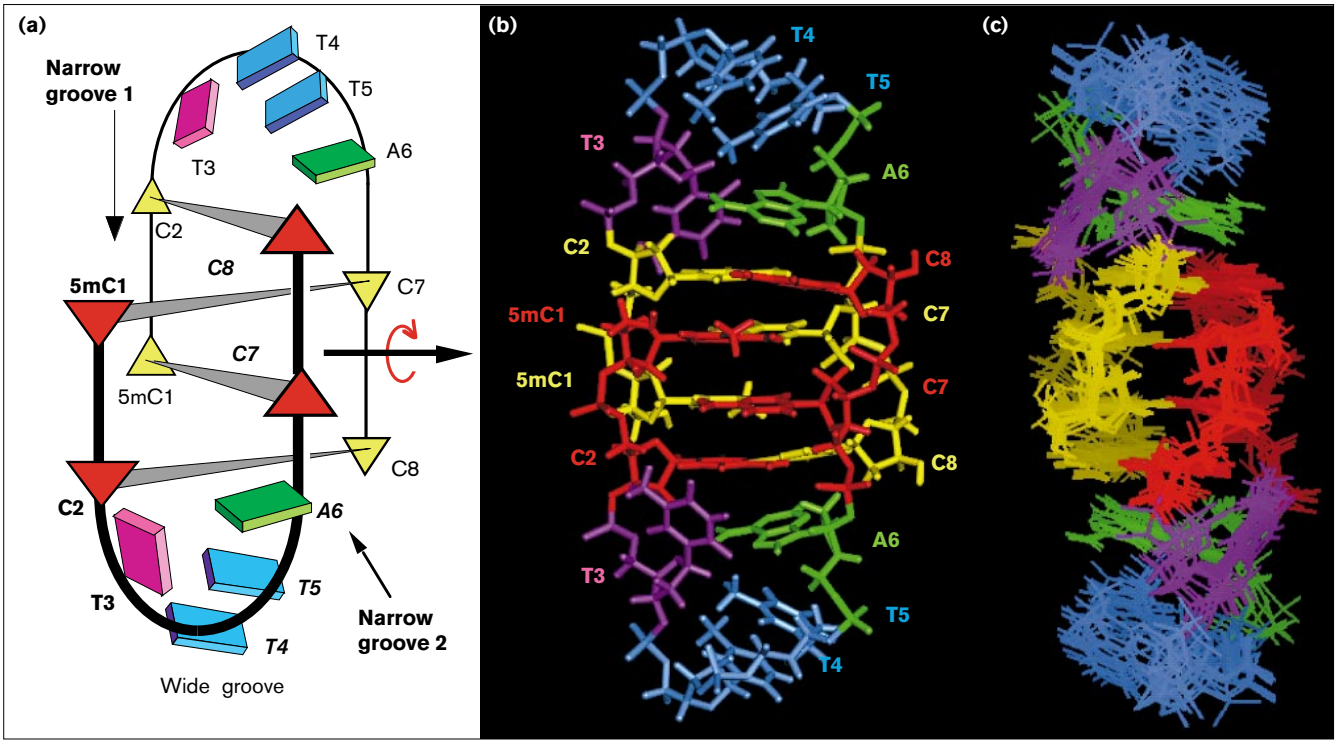
*Average and root mean square deviations (rmsds) are computed for 14 conformers. †The buckle is computed as the complement of the angle between the C6–N3 vector and the longitudinal symmetry axis; it is negative when the vector points in the 3′ direction. ‡The propeller twist is the complement of the angle between the C4–C2 vector and the longitudinal symmetry axis; it is negative when the vector points in the 3′ direction.

Both chemical exchange and dipolar interactions may contribute to the cross-peaks between amino protons and water. Moreover, H_2O–C2 $H4_{int}$ and H_2O–C8 $H4_{int}$ cross-peaks may include a dipolar contribution relayed by the fast exchanging imino proton of C2·C8$^+$.

[d(5mCCCTTTACCC)]$_2$

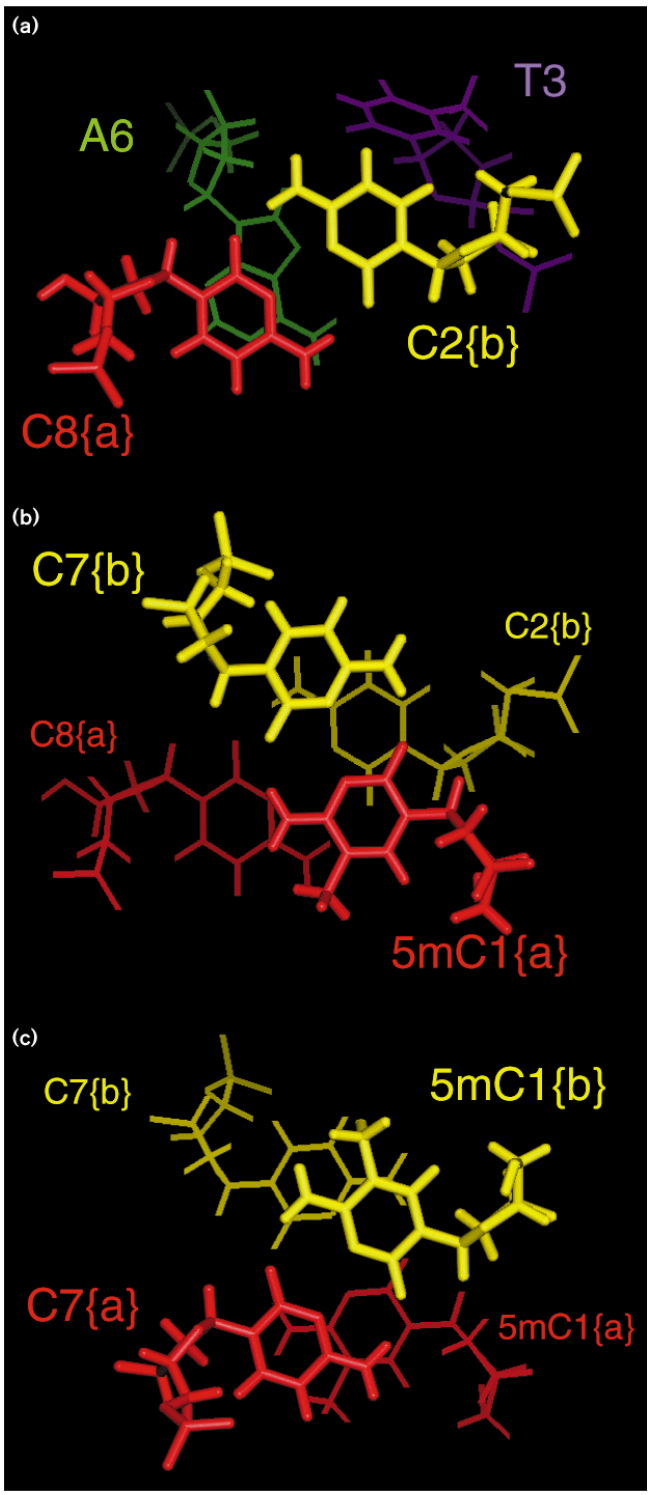
Gel filtration chromatography experiments performed on an aliquot of the NMR sample (concentration ~5 mM) showed that the structure formed by d(5mCCCTTTACCC) migrates as a dimer. The NOESY cross-peaks between

Figure 9



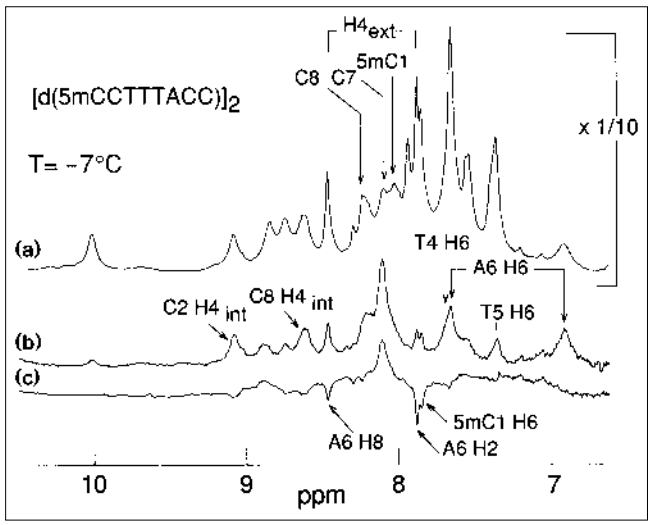
Structure of the [d(5mCCTTTACC)]$_2$ i-motif. **(a)** Schematic representation of [d(5mCCTTTACC)]$_2$. The cytidine color-coding is the same as in the schematic representation in Figure 5b, A6 is in green, T3 is in purple, and T4 and T5 are in blue. The C2 symmetry axis passes through the two non-equivalent narrow grooves. **(b)** View of the lowest-energy conformer looking into the wide groove. **(c)** View of a superposition of the eight lowest-energy conformers of [d(5mCCTTTACC)]$_2$, looking into the narrow groove defined by the 5mC1–C2 segments.

Figure 10



Base stacking in [d(5mCCTTTACC)]$_2$. The same color-coding is used as in Figure 9. **(a)** The adenosine of strand {b} is stacked to C8{a} and T3 is flipped out into the wide groove. **(b)** Stacking of 5mC1·C7$^+$ onto C8·C2$^+$. The groove between residues C7 and C8 is broader than between 5mC1 and C2. **(c)** Stacking of the two 5mC1·C7$^+$ pairs.

Figure 11



NOEs from water to aromatic and amino protons of [(d5mCCTTTACC)]$_2$. **(a)** Reference spectrum. **(b)** 1D-NOE detected 70 ms after selective inversion of the water magnetization. **(c)** 1D-ROE spectrum obtained with a 30 ms spin-lock. For normalization, the spectra were divided by the number of free induction decays (FIDs) recorded. The reference spectrum is multiplied by 1/10 for comparison with spectra (b) and (c). The sign of the NOEs and ROEs detected between water and A8 H6, A6 H2 and 5mC1 H6 indicate the proximity of water molecules whose residence times are longer than 0.5 ns. Experimental conditions: T = −7°C, pH 4.2.

imino and amino protons, which indicated three pairs of non equivalent cytidines, established that [d(5mCCCTTTACCC)]$_2$ is formed of two loops in a head-to-tail orientation, as in the case of [d(5mCCTTTACC)]$_2$.

The NOESY spectra in H_2O and D_2O provide evidence for the formation of an intercalated structure whose stacking order is 3*·10 / 1·8* / 2*·9 / 2·9* / 1*·8 / 3·10*.

The exchange times of C3·C10$^+$ and 5mC1·C8$^+$ imino protons are similar to those of the imino protons of pairs C2·C8$^+$ and 5mC1·C7$^+$, respectively, in [d(5mCCTTTACC)]$_2$. The exchange time of the C2·C9$^+$ imino proton, 480 s at 0°C, was determined from a real-time exchange experiment.

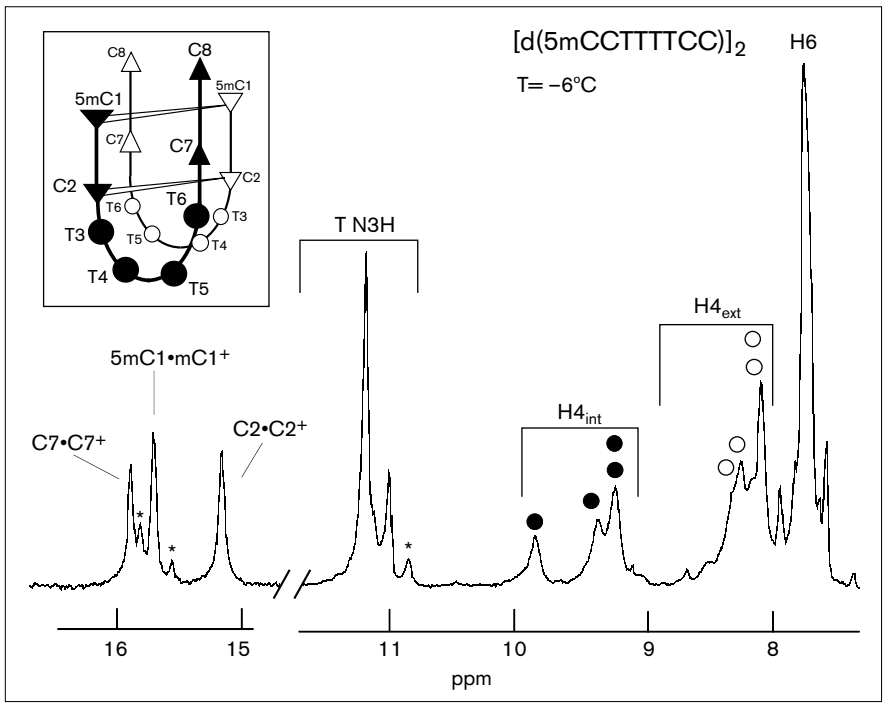
[d(5mCCTTTTCC)]$_2$

Gel filtration chromatography of an aliquot of the NMR sample (concentration 12 mM) shows that d(5mCCTTTTTCC) migrates as a dimer. The eight spin systems and the characteristic H1′–H1′ and amino–H2′/H2″ NOESY cross-peaks indicate a symmetrical i-motif dimer.

The exchangeable proton region shows four pairs of amino protons and three imino protons at positions expected for those of a C·C$^+$ pair (Figure 12). The missing imino proton

Figure 12

Proton spectrum and model (inset) of the i-motif dimer of [d(mCCTTTTCC)]$_2$. The observation that each imino proton is NOE-connected to a single amino proton group establishes the fact that the C·C$^+$ pairs are symmetrical. This requires that the four imino protons are aligned on a twofold axis and that the two TTTT loops are on the same side of the i-motif core. The NMR peaks marked by an asterisk originate from a minor species. Solution conditions: H_2O 90%, pH 4.6, T = −6°C, strand concentration 12 mM.



is presumably broadened out by fast exchange with water. The intensity of the cytidine imino protons is half that of the thymidines. The duplex symmetry revealed by the NOE connectivities between imino and amino protons is quite different from that described above for [d(5mCC-TTTACC)]$_2$. Each imino proton is connected to two equivalent amino groups. This indicates that the cytidines of the C·C$^+$ pairs are equivalent. Hence, each cytidine of one strand must pair with the same cytidine of the other strand. Taking into account the opposite orientations of the duplexes intercalated in the i-motif, only two intercalation orders are possible: 1*·1 / 8·8* / 2*·2 / 7·7* and 8*·8 / 1·1* / 7*·7 / 2·2*. In both cases, the TTTT loops are on the same side of the i-motif core (inset Figure 12).

Proton assignment began with the identification of the imino proton of the 5mC1·5mC1$^+$ pair (15.7 ppm) by means of the intraresidue connectivities between amino and methyl protons. As imino proton exchange from the external pair is expected to be fast, the observation that the imino proton of this pair exchanges only slowly favor the intercalation order with the C8·C8$^+$ pair in the terminal position of the i-motif stack, and suggests that the missing imino proton belongs to that pair. The 15.9 ppm imino proton, which is NOE-connected with the two other C imino protons (data not shown), is then assigned to C7·C7$^+$, the pair intercalated between pairs 5mC1·5mC1$^+$ and C2·C2$^+$, and the 15.2 ppm imino proton is therefore assigned to the C2·C2$^+$ pair. The four cytidine spin systems were assigned (except for H4′, H5′ and H5″ protons) from the TOCSY and the intraresidue NOE cross-peaks.

The identification of the cytidine spin systems thus obtained accounts for all the characteristic cross-peaks expected for the intercalation topology 8*·8 / 1·1* / 7*·7 / 2·2*, as shown in the inset of Figure 12. Three H1′–H1′ cross-peaks link C8 to 5mC1, 5mC1 to C7 and C7 to C2, and a reciprocal set of amino–H2′/H2″ cross-peaks connects C7 and C2.

In the model displayed in Figure 12, each C–C stretch (e.g. C8–C7) has two equivalent neighboring strands (5mC1*–C2* and 5mC1–C2 in this example), one across the wide groove and the other across the narrow groove. The symmetry of the structure prevents us from determining whether the NOEs between segments 5mC1–C2 and C8–C7 originate from intra- or interstrand connectivities. It is therefore impossible on the basis of the NMR data to determine if the TTTT loops bridge across the wide or the narrow groove.

Discussion

Loop folding

The NMR features related to the symmetry of the i-motif dimers provide crucial information about their folding topology. In [d(5mCCTTTACC)]$_2$ and [d(5mCC-TTTTCC)]$_2$, the NOE cross-peaks between imino and amino protons indicate immediately the symmetry (or

the asymmetry) of the hemiprotonated pairs and consequently the relative orientation of the loops.

The identification of the cytidine residues and the stacking order of the pairs were directly derived from the NMR spectra using the NOE connectivities characteristic of the i-motif. Conversely, there is no direct indication about the loop folding. As shown in Figure 5, the C·C$^+$ base pairs of $[d(5mCCTTTACC)]_2$ can be either intrastrand (in which case each TTTA sequence must follow a groove from the top to the bottom of the i-motif core) or interstrand (in which case the TTTA sequence loops across the wide grooves). Modeling performed without any distance constraints on the TTTA loop residues shows that both folding topologies are feasible. However, a model computed with intrastrand base pairs (Figure 5c) exhibits strong NOE violations when constrained by the distances derived from the NOE cross-peaks detected between T3 and C2, and between the C2·C8$^+$ pair and A6. This excludes the loop folding topology of the scheme shown in Figure 5c and leaves us with that of Figure 5b as a unique possibility.

Base pair opening kinetics of C·C$^+$ pairs

Imino proton exchange from C·C$^+$ pairs is controlled by the opening rate of the pairs [17]. The lifetime of 5mC1·C7$^+$ (1 s at 15°C) is comparable to that of the inner pair of $[d(TCC)]_4$ [13]. The lifetime of C2·C8$^+$ (15 ms at 0°C), which is about ten times longer than that of the outer pair of $[d(TCC)]_4$, reflects the contribution of the loop residues to the stability of this pair.

Hydrogen bonding in the TTTA loop

Exchange catalysis by hydroxide (Figure 7c) and phosphate (Figure 7d) indicates good accessibility of the T3, T4 and T5 imino protons. Imino proton exchange in monomeric thymidine is controlled by direct transfer to water between pH 3 and pH 5. In this pH range, T4 and T5 imino protons exchange at rates comparable to that of the monomer imino proton. In the case of T3, exchange is much faster (Figure 7c). This observation indicates the presence of a proton acceptor belonging to the nucleic acid itself. This feature is reminiscent of the intrinsic catalysis of the imino proton of terminal base pairs by the cyclic nitrogen of the complementary nucleoside, a process occurring from the open base pair through a concerted proton transfer across a water molecule [23,24]. T3 imino proton exchange may be catalyzed by nitrogens N1 or N7 of A6. In the computed structure, the A6 N7–T3 H3 distance (4.6 ± 0.5 Å) allows the insertion of the water molecule required for a concerted transfer. In this case, the exchange rate should be proportional to $10(pK_{(N7;A)}-pK_{(N3;T)})$. We measure (on Figure 7) that T3 imino proton exchange is five times faster than from the monomer. Assuming that the A6 N7–H_2O–H3 T3 water bridge is always formed, this would require $pK_{(N7;A)}$ to be equal to pK_{H_2O} + log(5) (i.e. –1). This value is not experimentally accessible because adenosine protonation occurs at the N1 position ($pK_{(N1;A)}$ = 4.05 at 0°C).

Due to its higher p*K*, nitrogen N1 is a better catalyst than N7. N1 is located far from T3 H3 in the computed structure, but transient rotation of the A6 ring around the glycosidic bond would bring N1 into a suitable orientation to catalyze exchange of the T3 imino proton. The observation that the exchange rate of T3 H3 is pH-independent around $pK_{(N1;A)}$ is, nevertheless, not in favor of a process catalyzed by N1.

In the i-motif structure formed by the intramolecular folding of d(5mCCTTTCCT8TTA11CCTTTCC), the TTTA loop also spans the wide groove and its structure is comparable to that of the loop of $[d(5mCCTTTACC)]_2$ except that T8 H3 is directly hydrogen bonded to A11 N7 (X Han *et al.*, unpublished data).

The detection of a water molecule close to A6 H8 in $[d(5mCCTTTACC)]_2$ (Figure 11) supports the existence of a water bridge between T3 H3 and A6 N7. However, similar NOE and ROE cross-peaks are observed between A11 H8 and water in d(5mCCTTTCCT8TTA11CCTTTCC), in a case where the T8 H3–A11 N7 distance is too short to accommodate the insertion of a water molecule.

Hence, we conclude that the fast exchange rate measured for the T3 imino proton in $[d(5mCCTTTACC)]_2$ indicates hydrogen bonding via a water molecule to a proton acceptor (probably A6 N7), but the identification of this water bridge to the long lived water molecule detected close to A6 H8 may be deceptive.

In the structure of $[d(5mCCTTTACC)]_2$, the amino protons of A6 are close to T3 O4′ (2.8 Å) and C2 O2 (3.7 Å). These distances are consistent with the presence of water bridges which could be responsible for the hindered rotation of the amino group.

Structure of the TTTA loop

The i-motif core is extended by the stacking of A6 onto C8 (stacking interval 2.7 Å; Figure 10). The sequential NOEs detected along the A6/T5/T4 track show a regular arrangement of the corresponding residues. The T4 and T5 rings are stacked together (stacking interval 3.2 Å). The lack of NOE cross-peaks between T3 and T4 indicates a sharp turn of the sugar phosphate backbone between these residues. The turn is characterized in particular by anomalous dihedral angle values for T3 ε (*gauche*$^-$), T3 ζ (*trans*) and T4 γ (*gauche*$^+$). The T3 base loops out in the major groove between C2·C8$^+$ and the A6 ring (Figures 9 and 10). It is tilted by about 60° with respect to that of the C8 plane.

It is remarkable that substitution of A6 by T6 changes the folding topology of the structure. This emphasizes the

stacking contribution of A6 to the stability of [d(5mCCTTTACC)]$_2$, and shows how the folding of intramolecular i-motifs may be guided by the composition of the linkers which connect the cytidine stretches.

Comparison with other DNA structures

The i-motif dimer of d(5mCCTTTACC) has a folding topology similar to that of the dimeric G tetrads formed by d(GCGGTTTGCGG) [25] and by d(GGGGTTTTGGGG) in the crystal structure [1]. All these dimers are formed by the association of two equivalent hairpins in a head-to-tail disposition and possess a twofold symmetry axis normal to one groove.

Geometry of the i-motif core

The i-motif dimer of [d(5mCCTTTACC)]$_2$, does not differ greatly from that of the i-motif tetramers studied so far with regard to helical twist, sugar conformation and glycosidic angles. Intercalation stretches the DNA phosphate backbone of all the intercalated residues (C1, C2, C7, C8 and also A6), leading to N sugar puckering. The shortest phosphorus–phosphorus distances across the narrow groove, between the 5mC1–C2 stretches (C2[P]–C2[P] = 7.6 ± 0.6 Å) and across the two equivalent wide grooves (C2[P]–C7[P] = 14.2 ± 0.6 Å), fall in the range commonly observed in i-motif structures, but the narrow groove between the C7–C8 stretches is unusually broad (C8[P]–C8[P] = 10.6 ± 1 Å; Figure 10).

Folding of the TTTA loop

Stabilization of the i-motif structure via stacking interactions with adenosine has been reported for the tetramers of d(CCCTAA) [12] and of d(TAACCC) [11]. Tetra loops in DNA and RNA hairpins fold into a limited number of well-defined conformations, classified by Hilbers *et al.* [26]. It is of interest to note that the structure of the TTTA loop of the i-motif dimer is different from that observed in the B-DNA hairpin of d(ATCCTAT7T8T9A10TAGGAT) [27]. In the B-DNA hairpin, T7 and A10 form a Hoogsteen base pair stacked on the last base pair of the stem, T9 stacks on T7·A10 and T8 is flipped towards the minor groove. The backbone makes a sharp turn between T9 and A10. The differences in TTTA loop structures do not arise from a difference in stem widths as the widths are comparable in the B-DNA hairpin (14–18 Å) and in [d(5mCCTTTACC)]$_2$ (14–15 Å). The stretching of the C7–A6 step, which is required in the i-motif dimer by the intercalation of the C2·C8$^+$ pair, might be at the origin of the differences.

Biological implications

DNA molecules can adopt many noncanonical structures, including antiparallel or parallel duplexes, triple helices and four-stranded structures. Among these structures, the so-called 'i-motif' has the unique property to be formed by the intercalation of two parallel duplexes in a head-to-tail orientation. Base pairing allows cytidine protonation at pH values much higher than the cytidine pK (p$K_{(cyt; N3)}$ = 4.3), in a range close to physiological pH.

Telomeres are nucleoproteic complexes localized at the ends of eukaryotic chromosomes. Telomeric DNA is composed of one strand containing short cytosine-rich repeats, together with a complementary guanosine-rich strand which protrudes in the 3′-direction and forms an overhanging single strand of several repeating units. These sequences play a role in maintaining the integrity and stability of the chromosome during replication. The C-rich composition of the telomeric DNA enables the formation of an i-motif structure.

X-ray studies have demonstrated that the repeat of the vertebrate telomere, either d(TAACCC) or d(CCCTAA), associates into an i-motif tetramer [11,12]. The d[(CCCTAA)$_3$CCC] oligonucleotide was shown by NMR to fold into a unimolecular i-motif at ~pH 7 [6]. It is noteworthy that the complementary G-rich sequence, d[GGG(TTAGGG)$_3$], folds *in vitro* into an intramolecular G quartet [2]. This leads to the enticing possibility that both strands of the chromosomal DNA could fold into four-stranded structures.

The centromeres are the nucleoproteic complexes which hold together the sister chromatids during meiosis and mitosis. The observation that stretches of only two cytidines, such as d(TCC) or d(CCTTTACC), associate into an i-motif raises the possibility that the d(CCATT) repeats of the human centromere could also adopt the i-motif structure. As in the case of telomeres, unusual structures have been proposed for the complementary centromeric G-rich strand [28,29].

The NMR spectra of oligonucleotides containing four telomeric repeats from vertebrates, from *Tetrahymena* [6], from *Saccharomyces cerevisiae* (J-LL, unpublished data) and *Bombix mori*, (X Han, unpublished data) and the spectrum of the oligonucleotide containing four human centromeric repeats (SN and J-LL, unpublished data) give indications for the formation of i-motif structures at neutral or slightly acidic pH. However, the poor quality of these NMR spectra, which probably result from conformational exchange between multiple species, has until now thwarted structural investigations. In the work reported here, the structure of the natural sequences has been approached by the study of oligonucleotides of similar composition, but whose folding is unique.

The observation that G- or C-rich strands can dimerize *in vitro* may be relevant to chromosomal pairing during meiosis [30–33]. Several studies have argued in favor of a biological role for G tetrads, and it has been suggested that they could be involved in the regulation of telomerase

activity [4]. Proteins have been shown to bind to G tetrads [34–36] or to promote their formation [37,38]. Evidence for a protein that binds specifically to the vertebrate telomeric single strand d(CCCTAA)n has been reported [39], but proteins which bind specifically to the i-motif have not yet been discovered. This may be because the discovery of the i-motif is more recent (1993) than that of the G tetrads (1988) and it would be premature to rule out the existence of i-motif specific proteins and the possibility of a biological function for the i-motif.

Materials and methods

Oligomer synthesis and purification

The DNA oligomers were synthesized and purified as previously described [13]. The strand concentration of the NMR samples was determined from the absorbance at pH 7 using A_{260} values of 63 400 M^{-1} and 57 799 M^{-1} for d(5mCCTTTACC) and d(5mCCTTTTCC), respectively, as computed with a nearest neighbor model [40].

NMR samples

Unless otherwise stated, the experiments were performed at a pH close to 4.3, the cytidine $pK_{(N3;\,C)}$. At this pH, the fraction of protonated cytidines in the hemiprotonated dimer and in the single strand is the same. Hence the solution pH, buffered by the cytidine residues, does not change upon dimer dissociation.

Gel filtration chromatography

The stoichiometry of the oligomers was determined by high pressure chromatography using a Synchropack GPC 100 gel filtration column calibrated with mononucleosides and oligonucleosides as previously described [6].

NMR methods

The NMR experiments were performed either on a 360 MHz home-built spectrometer or a 600 MHz Brucker AMX spectrometer. Experiments in water used a 'Jump and Return' sequence [41] for detection with the maximum sensitivity set at 13.5 ppm. Water signal suppression was improved using standard NMR tubes welded to an 8 mm long Pyrex tail.

In two-dimensional (2D) experiments, the number of free induction decays accumulated for each t1 value was proportional to the value at time t1 of the apodization function used for data processing in that dimension. The quadrature detection was accomplished using the States method [42] . NOESY experiments used 90i pulses for preparation and mixing. For NOESY experiments performed in H_2O, a 1 ms Z gradient was applied at the beginning of the mixing period and the signal was detected with a JR sequence. The spectral width was 8.5 kHz, the acquisition time 0.12 s, and the repetition rate 1.5 s. The t1 delay was incremented from 0 to 60 ms in 240 steps. NOESY and TOCSY experiments in D_2O were recorded at 7°C. The spectral width was 3.5 kHz, the acquisition time 0.29 s, and the repetition rate 1.6 s. The t1 delay was incremented from 0 to 84 ms (256 increments). Generally, the residual HDO signal was not saturated in order to avoid saturation of the H3′ protons. TOCSY experiments used 10 MLEV-17 repetitions (total time 15 ms) without trim pulses [43].

1D NOE and ROE experiments designed to detect hydration water molecules [44] were recorded as a function of the mixing times after water labeling (5 to 400 ms) at different temperatures (−7, 10, 25 and 40°C). A spin-echo filter [45], based on the T2 difference between H_2O and the H3′ protons, was used to cancel the magnetization of the nucleic acid protons at the water frequency. The spin-echo filter delay ranged typically from 80 ms at −7°C to 200 ms at 40°C. In 1D ROE experiments, the water magnetization was continuously spin-locked by a DANTE sequence [(π/4), 20 ms)]n [46,47]. To ensure effective water suppression, the water magnetization was flipped back to the B_0 direction by a 90° pulse followed by a 0.5 ms Z gradient pulse (0.076 T/m) before JR detection. All 1D-NOE and ROE spectra were obtained with a repetition time of 3 s.

Data processing

1D spectra were corrected for the frequency response to the JR excitation and the residual water signal was reduced by post-acquisition processing in the time domain [48]. In NOESY experiments in H_2O, the residual water signal was reduced by a digital shift correction [49]. The 2D data were processed on an Indy work station (Silicon Graphics Inc.) with Felix 95.1 software (Biosym). NOESY in H_2O were zero-filled to 1536 complex points in both dimensions, and experiments in D_2O were zero-filled to 1024 points in dimension t1. The spectra were apodized with 3 Hz exponential broadening and a 45° phase shifted square sine bell function in both dimensions.

Proton exchange theory

Imino proton exchange from the thymidine monomer is catalyzed by added proton acceptors such as phosphate [50,51]. In the absence of added proton acceptor, exchange is controlled by hydroxyl catalysis above pH 5. Imino proton exchange controlled by direct transfer to water is pH-independent between pH 5 and pH 3. It occurs at a rate depending on the pK difference between the thymidine N3 ($pK_{(N3;\,T)} = 10$ at 0°C) and the H_2O/H_3O^+ couple, ($pK_{H_2O} = -1.7$) [52]: kex, $H_2O \sim 10(pK_{H_2O} - pK_{(N3;\,T)})$.

Exchange of the hydrogen bonded imino proton cannot occur from the closed base pair. It requires opening of the pair followed by proton transfer to an acceptor acting as a catalyst. In the open base pair, the transfer proceeds at a rate comparable to that of the monomer.

Imino proton exchange from the open pair may also be catalyzed by a proton acceptor belonging to the oligonucleoside itself. The cyclic nitrogen of the paired nucleotide (A N1 in Watson–Crick A·T pairs; C N3 in $C{\cdot}C^+$ pairs) acts as an intrinsic catalyst in a concerted transfer involving a water molecule [23]. In the case of $C{\cdot}C^+$ base pairs, the intrinsic catalysis is so efficient that the imino proton exchanges at each opening event. The exchange time is then identical to the base pair lifetime [17].

Exchange time measurements

The methods for proton exchange time measurements have been reviewed recently [51]. In the present work, imino proton exchange rates were derived from the rate of magnetization transfer from water after selective inversion of the water magnetization using a DANTE sequence followed by a 1 ms Z gradient pulse to cancel out the residual transverse component. Exchange times in the millisecond range were determined from the line- broadening induced by proton exchange.

Structure determination

NMR determinants of the i-motif

Intercalation of two duplexes into an i-motif results in large distances between sequentially adjacent residues, and in the quasi-absence of sequential NOE connectivities at short mixing times. The intercalation topology of the i-motif is read off the patterns of H1′–H1′ cross-peaks which connect the stacked residues across the narrow grooves [9] and off the amino–H2′/H2″ cross-peaks connecting across the wide grooves the nucleotides stacked via contacting 3′-faces. Stacked $C{\cdot}C^+$ pairs are also connected by strong H5–amino proton cross-peaks [13]. The intercalation stretches the sugar phosphate backbones, leading to N-type puckers and strong intraresidue H3′–H6 cross-peaks.

NOESY cross-peaks and distance constraints

The structure was determined from the build-up rates of 35 interresidue and 29 intraresidue NOE cross-peaks measured at mixing times of 50, 70, 90, 120, 160, 200 ms in D_2O and of 45, 67, 90, 120, and 200 ms in H_2O at 7°C.

The NOESY cross-peak volumes were defined manually and measured with Felix 95.1 software. They were corrected as needed for the

frequency response to JR excitation and for the digital-shift procedure. The volumes were converted to distances by comparison of the initial slope of their built-up rate versus the mixing time with that of the intraresidue cytidine H5–H6 cross-peaks, using a reference distance of 2.45 Å for the latter. Distances involving a methyl group were scaled by reference to the H6–CH_3 pseudo-atom distance of 2.9 Å.

The distances were sorted into three categories corresponding to lower and upper bounds of 1.8–2.9, 1.8–3.7 and 2.9–4.2 Å. In case of the cross-peaks to geminal protons, the distance corresponding to the strongest cross-peak was used as constraint and those derived from the other cross-peaks were considered as lower bounds. For distances involving methyl protons or C·C^+ imino protons, the bounds were loosened by ± 0.5 Å. Cross-peaks detected at long mixing times, and in some cases the absence of cross-peaks, were translated into repulsive constraints with a lower bound of 4.2 Å. All the distance constraints were duplicated to take into account the dimer symmetry.

Base-pairing constraints

The two strands of the dimer were designated {a} and {b}. The C·C^+ base pairs were defined by the O2–$H4_{int}$ and N3–N3 distances which were set to 1.74 ± 0.1 Å and 2.76 ± 0.1 Å, respectively. The base-pair planarity was not enforced.

Strand equivalence and symmetry constraints

According to the NMR spectrum, the two strands were declared as equivalent groups by the 'noncrystallography symmetry protocol' of X-PLOR [53]. In order to enforce the symmetry between strands in the dimer, several equivalent {a}–{b} and {b}–{a} interstrand distances were declared identical within 0.1 Å [13]. Using this procedure, the rmsd between the two strands of a conformer was much smaller than between two computed conformers.

As a consequence of the symmetry of the structure, the assignment of the NOE cross-peaks to intra- or interstrand connectivities is ambiguous. To solve this ambiguity, we built a preliminary model constrained by the interresidue distances derived from the characteristic amino–H2′/H2″ and H1′–H1′ cross-peaks, and by the base-pairing constraints. Using this model, we assigned the other cross-peaks to the shortest of the homolog intra- or interstrand distances. Another model built using the SUM averaging procedure [18] of the X-PLOR program with ambiguously specified distance constraints, yielded the same assignments of cross-peaks to intra or interstrand distances.

Distance-restrained molecular dynamics

Restrained molecular dynamics was carried out on a Hewlett Packard Apollo 715/50, using the simulated-annealing protocol of the X-PLOR program [53] with the standard harmonic potential for covalent geometry. Potential energy terms related to electrostatics and 'empirical dihedral' were omitted. The force constants related to NOE-derived distance constraints and base-pairing constraints were set to 50 kcal mol^{-1} $Å^{-2}$ and 500 kcal mol^{-1} $Å^2$, respectively. The NMR symmetric constraints were maintained throughout. The computation started with two extended d(5mCCTTTACC) strands in an arbitrary conformation. At the first step of the modeling procedure, the energy of the initial structure was minimized by five Powell cycles. The hydrogen bond force constant was turned off at the beginning of the dynamics. The initial velocity was set to correspond to 1200K, and the molecular dynamics computation was run for 1000 steps of 2 ps. The hydrogen-bonding energy was introduced during the procedure, which reduced the temperature to 300K in steps of 25K, each step being followed by a 0.1 ps dynamics computation. The energy of the resulting conformers was minimized by 300 Powell cycles. The computation was repeated to yield 50 conformers among which 14 were selected for their low NOESY-related energy. The rmsd were determined by pair-wise comparison of the conformers aligned on the heavy atoms of the cytidine rings. The conformers computed using an initial velocity corresponding to 2000K and a dynamic computation of 2000 steps of 2 ps were not significantly different. We are aware that short intraresidue distance constraints may be spoiled by spin diffusion. In order to test the sensitivity of the structure to such effects, we computed a set of conformers without intraresidue constraints. These conformers have the same intercalation topologies, loop folding, glycosidic torsion angles, and N or S sugar puckering as those computed with the intraresidue distance constraints.

Accession numbers

The coordinates of the lowest NOESY-related energy conformer $[d(5mCCTTTACC)]_2$ have been submitted to the Protein Data Bank (accession code 1BAE) and are available from the authors on request until released.

Acknowledgements

We thank Maurice Guéron for stimulating discussions and help in the preparation of the manuscript. Monique Leblond is acknowledged for help in the oligomer synthesis and purification. This work was supported by Grant 6520/6 July 1994 from the Association pour la Recherche contre le Cancer. Sylvie Nonin was partially supported by the Fondation pour la Recherche Medicale.

References

1. Kang, C., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. (1992). Crystal structure of four-stranded *Oxytricha* telomeric DNA. *Nature* **356**, 126–131.
2. Wang, Y. & Patel, D.J. (1993). Solution structure of the human telomeric repeat d[AG3(T2AG3)$_3$] G-tetraplex. *Structure* **1**, 263–282.
3. Smith, F.W., Schultze, P. & Feigon, J. (1995). Solution structures of unimolecular quadruplexes formed by oligonucleotides containing *Oxytrichia* telomere repeats. *Structure* **3**, 997–1008.
4. Zahler, A.M., Williamson, J.R., Cech, T.R. & Prescott, D.M. (1991). Inhibition of telomerase by G-quartet DNA structure. *Nature* **350**, 718–720.
5. Ahmed, S., Kintanar, A. & Henderson, E. (1994). Human telomeric C-strand tetraplexes. *Nat. Struct. Biol.* **1**, 83–88.
6. Leroy, J.L., Guéron, M., Mergny, J.L. & Hélène, C. (1994). Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. *Nucleic Acids Res.* **22**, 1600–1606.
7. Langridge, R. & Rich, A. (1963). Molecular structure of helical polycytidylic acid. *Nature* **198**, 725–728.
8. Hartman, K.A., Jr. & Rich, A.J. (1965). The tautomeric form of helical polyribocytidylic acid. *J. Am. Chem. Soc.* **87**, 2033–2039.
9. Gehring, K., Leroy, J.L. & Guéron, M. (1993). A tetrameric DNA structure with protonated cytidine–cytidine base pairs. *Nature* **363**, 561–565.
10. Chen, L., Cai, L., Zhang, X. & Rich, A. (1994). Crystal structure of a four-stranded intercalated DNA: d(C_4). *Biochemistry* **33**, 13540–13546.
11. Kang, C., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. & Rich, A. (1995). Stable loop in the crystal structure of the intercalated four-stranded cytosine-rich metazoan telomer. *Proc. Natl. Acad. Sci. USA* **92**, 3874–3878.
12. Berger, I., Kang, C., Fredian, A., Ratliff, R., Moyziz, R. & Rich, A. (1995). Extension of the four-stranded intercalated cytosine motif by adenine–adenine base-pairing in the crystal structure of d(CCCAAT). *Nat. Struct. Biol.* **2**, 416–425.
13. Leroy, J.L. & Guéron, M. (1995). Solution structure of the i-motif tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): novel NOE connection between amino protons and sugar protons. *Structure* **3**, 101–120.
14. Rohozinski, J., Hancock, J.M. & Keniry, M.A. (1994). Oligocytidine regions contained in DNA hairpin loops interact via a four-stranded, parallel structure similar to the i-motif. *Nucleic Acids Res.* **22**, 4653–4659.
15. Mergny, J.L., Lacroix, L., Han, X., Leroy, J.L. & Hélène, C. (1995). Intramolecular folding of pyrimidine oligodeoxynucleotides into an i-motif. *J. Am. Chem. Soc.* **117**, 8887–8898.
16. Nonin, S. & Leroy, J.L. (1996). Structure and conversion kinetics of a bi-stable DNA i-motif: broken symmetry in the $[d(5mCCTCC)]_4$ tetramer. *J. Mol. Biol.* **261**, 399–414.
17. Leroy, J.L., Gehring, K., Kettani, A. & Guéron, M. (1993). Acid multimers of oligo-cytidine strands: stoichiometry, base-pair characterization and proton exchange properties. *Biochemistry* **3**, 6019–6031.

18. Nilges, M. (1995). Calculation of protein structures with ambiguous distance restraints; automated assignment of ambiguous NOE cross peaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660.
19 Otting, G., Liepinsh, E. & Wüthrich, K. (1991). Protein hydration in aquous solution. *Science* **254**, 974–980.
20. Liepinsh, E., Otting, G. & Wütrich, K. (1992). NMR observation of individual molecules of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution. *Nucleic Acids Res.* **20**, 6549–6553.
21. Kubinec, M.G. & Wemmer, D. (1992). NMR evidence for DNA bound water in solution. *J. Am. Chem. Soc.* **114**, 8739–8740.
22. Zhou, D. & Bryant, R.G. (1996). Water molecule binding and lifetimes on the DNA duplex d(CGCGAATTCGCG)$_2$. *J. Biomol. NMR* **8**, 77–86.`
23. Guéron, M., Kochoyan, M. & Leroy, J.L. (1987). A single mode of DNA base-pair opening drives imino proton exchange. *Nature* **382**, 89–92.
24. Nonin, S., Leroy, J.L. & Guéron, M. (1995). Terminal base pair of oligodeoxynucleotides: imino proton exchange and fraying. *Biochemistry* **34**, 10652–10659.
25. Kettani, A., Kumar, A. & Patel, D.J. (1995). Solution structure of a DNA quadruplet containing the fragile X syndrome triplet repeat. *J. Mol. Biol.* **254**, 638–656.
26. Hilbers, C.W., Heus, H.A., van Dongen, M.J.P. & Wijmenga, S.S. (1994). The hairpin elements of nucleic acid structure: DNA and RNA folding. *Nucleic Acids Mol. Biol.* **8**, 56–104.
27. Blommers, M.J.J., van de Ven, F.J.M., van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1991). The three-dimensional structure of a DNA hairpin in solution. Two dimensional NMR studies and structural analysis of d(ATCCTATTTATAGGAT). *Eur. J. Biochem.* **201**, 33–51.
28. Chou, S.H., Cheng, J-W., Fedoroff, O. & Reid, B.R. (1994). DNA sequence GCGAATGAGC containing the human centromere core sequence GAAT forms a self-complementary duplex with sheared G·A pairs in solution. *J. Mol. Biol.* **241**, 467–479.
29. Gupta, G., Garcia, A.E., Castati, P., Ratliff, R., Bradbury, E.M. & Moyzis, R.K. (1993). Stem-loop structures of the repetitive DNA sequences located at the human centromeres. In *Structural Biology: the State of the Art. Proceedings of the Eight Conversation, State University of New York, Albany, NY, USA*. (Sarma, R.H. & Sarma, M.H., eds), pp. 137–154, Vol 2, Adenine Press, Schenectady, NY.
30. Sen, D. & Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364–366.
31. Ashley, T. & Ward, D.C. (1993). A hot-spot of recombination coincides with an interstitial telomeric sequence in the Armenian hamster. *Cytogenet. Cell Genet.* **62**, 169–171.
32. Katinka, M.D. & Bourgain, F.M. (1992). Interstitial telomers are hot spots for illegitimate recombination with DNA molecules injected into the macronucleus of *Paramecium primaurelia*. *EMBO J.* **11**, 725–732.
33. Pluta, A.F. & Zakian, V.A. (1989). Recombination occurs during telomere formation in yeast. *Nature* **337**, 429–433.
34. Weisman-Shomer, P. & Fry, M. (1993). QUAD, a protein from hepatocyte domain chromatin that binds selectively to guanine-rich quadruplex DNA. *J. Biol. Chem.* **268**, 3306–3312.
35. Schierer, T. & Henderson, E. (1994). A protein from *Tetrahymena thermophila* that specifically binds parallel-stranded G4-DNA. *Biochemistry* **33**, 2240–2246.
36. Baran, N., Pucshansky, L., Marco, Y., Benjamin, S. & Manor, H. (1997). The SV40 large T-antigen helicase can unwind four-stranded DNA structures linked by G-quartets. *Nucleic Acids Res.* **25**, 297–303.
37. Giraldo, R. & Rhodes, D. (1994). The yeast telomer-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.* **13**, 2411–2420.
38. Giraldo, R., Suzuki, M., Chapman, L. & Rhodes, D. (1994). Promotion of parallel DNA quadruplexes by a yeast telomer binding protein: a circular dichroism study. *Proc. Natl. Acad. Sci. USA* **41**, 7658–7662.
39. Marsich, E., Piccini, A., Xodo, L.E. & Manzini, G. (1996). Evidence for a HeLa nuclear protein that binds specifically to the single stranded d(CCCTAA)n telomeric motif. *Nucleic Acids Res.* **24**, 4029–4033.
40. Cantor, C.R. & Warshaw, M.M. (1970). Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers* **9**, 1059–1077.
41. Plateau, P. & Guéron, M. (1982). Exchangeable proton NMR without base-line distortion. *J. Am. Chem. Soc.* **104**, 7310–7311.
42. States, D.J., Haberkorn, R.A. & Reuben, D.J. (1982). A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J. Magn. Reson.* **48**, 286–292.
43. Bax, A. & Davis, D.G. (1985). MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355–360.
44. Otting, G. & Liepinsh, E. (1995). Selective excitation of water signal by a Q-switched selective pulse. *J. Magn. Reson. B* **107**, 192–196.
45. Mori, S., Berg, J.M. & van Zijl, P.C.M. (1996). Separation of intramolecular NOE and exchange peaks in water exchange spectroscopy using spin-echo filters. *J. Biomol. NMR* **7**, 77–82.
46. Kessler, H., Greisinger, C., Kerssebaum, R., Wagner, K. & Ernst, R.R. (1987). Separation of cross-relaxation and J cross-peaks in 2D rotating-frame NMR spectroscopy. *J. Am. Chem. Soc.* **109**, 607–609.
47. Morris, G.A. & Freeman, R. (1978). Selective excitation techniques in Fourrier transform NMR. *J. Magn. Reson.* **29**, 433–462.
48. Guéron, M. Plateau, P. & Decorps, M. (1991). Solvent signal suppression in NMR. In *Progress in NMR Spectroscopy.* (Emsley, J.W., Feeney, J. & Sutcliffe, L.H., eds.), pp. 135–209, **23**, Pergamon Press, Oxford, UK.
49. Roth, K., Kimber, B.J. & Feeney, J. (1980). Data shift accumulation and alternate delay accumulation techniques for overcoming the dynamic range problem. *J. Magn. Reson.* **41**, 302–309.
50. Eigen, M. (1964). Proton transfer, acid-base catalysis, and enzymatic hydrolysis. Part I: elementary processes. *Angew. Chem.* **3**, 1–19.
51. Guéron, M. & Leroy, J.L. (1996). Studies of base pair kinetics by NMR measurement of proton exchange. *Methods Enzymol.* **261**, 383–413.
52. Nonin, S., Leroy, J.L. & Guéron, M. (1996). The acid-catalysed imino proton exchange of G·C Watson–Crick base pairs. *Nucleic Acids Res.* **24**, 586–595.
53. Brünger, A.T. (1990). X-PLOR Version 3.1, A system for X-ray crystallography. Yale University, New Haven, CT.